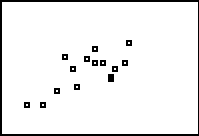
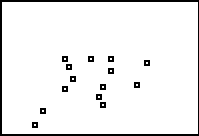
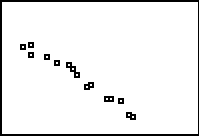
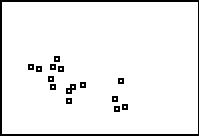
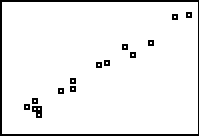
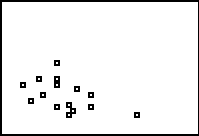
**Stat and Data Analysis Name: KEY**

**Worksheet 4.1A- Interpreting Scatterplots**

**Class A Class B Class C**

**  **

**Class D Class E Class F**

**  **

Consider the above scatterplots of hypothetical scores on the first and second exams of a course. The horizontal axis represents the first exam and the vertical axis represents the second exam. Each axis is formatted form 50 to 100 with a scale of 5.

Describe the scatterplots above (form, direction, strength), then fill in the table below with the appropriate letter of each graph

A: Linear, Moderate, Positive

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Strong** | **Moderate** | **Weak** |
| **Negative** | C | D | F |
| **Positive** | E | A | B |

B: Linear, Moderately Weak, Positive

C: Linear, Strong, Negative

D: Linear, Moderate, Negative

E: Linear, Strong, Positive

F: Linear, Weak, Negative, Outlier

For the pairs of variables below, indicate what you would expect for the direction (positive, negative, scattered) and the strength (none, weak, moderate, strong) of the association.

1. Height and armspan Positive, Strong
2. Height and shoe size Positive, Moderate
3. Height and GPA Scattered, None
4. SAT score and college GPA Positive, Moderate
5. Latitude and average January temperature of American cities Negative, Moderate
6. Lifespan and weekly cigarette consumption Negative, Moderate
7. Serving size and calories of fast food sandwiches Positive, Moderate
8. Air fare and distance to destination Positive, Moderate

**Worksheet 4.1B- Correlation NEED: Program CORR, group HYPOCORR**

**Ungroup HYPOCORR (you should get lists EXA1, EXA2, EXB1, EXB2, etc.)**

**These lists are hypothetical exam scores, the same ones from Worksheet 7A**

1. Look back at the scatterplots from worksheet 7A to see the graphs for Classes A-F. Now use the lists that you have and compute the correlation for each graph, and fill in the correlation in the same table from before:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Strong** | **Moderate** | **Weak** |
| **Negative** | **C -0.9855** | **D -0.7198** | **F -0.4723** |
| **Positive** | **E 0.9888** | **A 0.7131** | **B 0.4653** |

1. Based on these results, what do you suspect is the largest value that a correlation coefficient can assume? What do you suspect is the smallest value?

**Largest: r = 1 Smallest: r = -1**

1. What types of scatterplots have the largest correlation coefficient? Which have the smallest?

**Strong positive linear associations have the largest correlation coefficients.**

**Strong negative linear associations have the smallest correlation coefficients.**

1. How does the value of the correlation relate to the *direction* of the association?

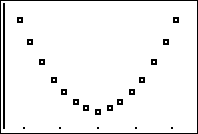
**+r = positive direction**

**-r = negative direction**

1. How does the value of the correlation relate to the *strength* of the association?

**Closer to |1| the stronger the linear association**

**Closer to 0 the weaker the linear association**

1. Make a scatterplot relating the scores from Exam G (EXG1 = x-list, EXG2 = y-list). Draw the graph. Does there seem to be any relationship between the scores? **Describe** the scatterplot.

**Curved**

**Strong association**

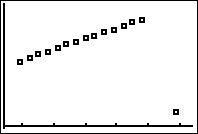
**Negative direction then switches to positive direction**

1. Calculate the correlation for Class G. Does its value surprise you? What type of relationship does *r* REALLY measure??

**r = 0; Correlation only measures linear relationships.**

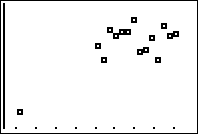
1. Calculate the correlation for Class G but switch the X and Y list. What is it? How does it compare to the original correlation? What does this tell you?

**r = 0; correlation doesn’t care which is the explanatory and which is the response variable**

1. Make a scatterplot for Class H. Sketch it below. Do most of the observations seem to follow a linear pattern? Are there any exceptions?

**All the points seem to fall on the line except for the outlier in the bottom right corner.**

1. Make a scatterplot for Class I. Do most of the observations seem to be scattered? Are there any exceptions?



**Most of the points are scattered.**

**There is one large outlier in the bottom left corner.**

1. Calculate the correlation for Classes H and I. Write them both down. Do either of these correlations surprise you? Why?

**Class H: *r* = 0.0365 Class I: *r* = 0.7046**

**Class H’s seems too small and Class I’s seems too large.**

1. Remove the outlier for Class H (be sure to remove the coordinates from both the X and Y lists). Recalculate the correlation. How has it changed?

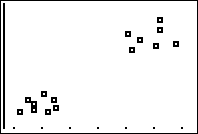
**Class H: *r* = 0.9997**

1. Remove the outlier for Class I. Recalculate the correlation. How has it changed?

**Class I: *r* = 0.1304**

1. Based on your analysis of Classes H and I, would you say that the correlation coefficient is *resistant* measure of association? Or *non-resistant*?

**Non-resistant! Outliers affect the correlation.**

1. Make a scatterplot for Class J. Describe what the plot reveals about the relationship between exam scores (describe the plot).

**Two clusters of observations. Neither shows much of an association.**

1. Calculate the correlation for Class J. Does this value surprise you? How?

***r* = 0.9544; It seems too high.**